

STATISTICS (STAT)

STAT 101 Introductory Business Statistics

Data summaries and descriptive statistics; introduction to a statistical computer package; Probability: distributions, expectation, variance, covariance, portfolios, central limit theorem; statistical inference of univariate data; Statistical inference for bivariate data: inference for intrinsically linear simple regression models. This course will have a business focus, but is not inappropriate for students in the college. One-term course offered either term
Prerequisite: MATH 104 or MATH 110 or equivalent; successful completion of STAT 101 is prerequisite to STAT 102
Activity: Lecture
1 Course Unit

STAT 102 Introductory Business Statistics

Continuation of STAT 101. A thorough treatment of multiple regression, model selection, analysis of variance, linear logistic regression; introduction to time series. Business applications. One-term course offered either term
Prerequisite: STAT 101
Activity: Lecture
1 Course Unit

STAT 111 Introductory Statistics

Introduction to concepts in probability. Basic statistical inference procedures of estimation, confidence intervals and hypothesis testing directed towards applications in science and medicine. The use of the JMP statistical package. One-term course offered either term
Prerequisites: High school algebra.
Activity: Recitation
1 Course Unit

STAT 112 Introductory Statistics

Further development of the material in STAT 111, in particular the analysis of variance, multiple regression, non-parametric procedures and the analysis of categorical data. Data analysis via statistical packages. One-term course offered either term
Prerequisite: STAT 111
Activity: Lecture
1 Course Unit

STAT 399 Independent Study

One-term course offered either term
Prerequisites: Written permission of instructor and the department course coordinator.
Activity: Independent Study
1 Course Unit

STAT 405 Statistical Computing with R

The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives. Taught by: Stine, Waterman, Zhang
One-term course offered either term
Prerequisite: STAT 102 or STAT 112 or STAT 430
Activity: Lecture
0.5 Course Units

STAT 422 Predictive Analytics for Business

This course follows from the introductory regression classes, STAT 102, STAT 112, and STAT 431 for undergraduates and STAT 613 for MBAs. It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest. One-term course offered either term
Prerequisite: STAT 102 or STAT 112 or STAT 431
Activity: Lecture
0.5 Course Units

STAT 424 Text Analytics

This course introduces methods for the analysis of unstructured data, focusing on statistical models for text. Techniques include those for sentiment analysis, topic models, and predictive analytics. Course includes topics from natural language processing (NLP), such as identifying parts of speech, parsing sentences (e.g., subject and predicate), and named entity recognition (people and places). Unsupervised techniques suited to feature creation provide variables suited to traditional statistical models (regression) and more recent approaches (regression trees). Examples that span the course illustrate the success of text analytics. Hierarchical generating models often associated with nonparametric Bayesian analysis supply theoretical foundations. Taught by: Stine
One-term course offered either term
Prerequisites: Students should be familiar with regression models at the level of STAT 102 and the R statistics language at the level of STAT 405. Familiarity with the R-Studio development environment is presumed, as well as common R packages such as stringr, dplyr and ggplot. Those with more knowledge of Statistics, such as from STAT 422, or computing skills will benefit. The predominant software used in the course is R, with bits of JMP when helpful for interactive illustration. Familiarity with basic probability models is helpful but not presumed.
Activity: Lecture
0.5 Course Units

STAT 430 Probability

Discrete and continuous sample spaces and probability; random variables, distributions, independence; expectation and generating functions; Markov chains and recurrence theory. One-term course offered either term
Prerequisite: MATH 114 or MATH 115 or equivalent
Activity: Lecture
1 Course Unit

STAT 431 Statistical Inference

Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course. This course does not have business applications but has significant overlap with STAT 101 and 102. One-term course offered either term
Prerequisite: STAT 430
Activity: Lecture
1 Course Unit

STAT 432 Mathematical Statistics

An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures.

Course usually offered in spring term

Prerequisite: STAT 430 or 510 or equivalent

Activity: Lecture

1 Course Unit

STAT 433 Stochastic Processes

An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes.

One-term course offered either term

Prerequisites: STAT 430, or permission of instructor

Activity: Lecture

1 Course Unit

STAT 435 Forecasting Methods for Management

This course provides an introduction to the wide range of techniques available for statistical forecasting. Qualitative techniques, smoothing and decomposition of time series, regression, adaptive methods, autoregressive-moving average modeling, and ARCH and GARCH formulations will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations.

Taught by: Shaman

One-term course offered either term

Prerequisite: STAT 102 or 112 or 431

Activity: Lecture

1 Course Unit

STAT 436 Introduction to Large-Scale Data Science

The course will focus on computational approaches to large-scale data analysis. The lectures will introduce the relevant concepts, and students will be asked to work on projects, implementing the methods and experimenting with large-scale datasets. The course will cover various techniques for updating models in an online fashion, as well as subsampling and dimensionality-reduction techniques. The students will experiment with neural network architectures and learn to build predictive models for modern machine learning tasks.

One-term course offered either term

Prerequisites: Linear Algebra and basic R programming

Activity: Lecture

1 Course Unit

STAT 442 Introduction to Bayesian Data Analysis

The course will introduce data analysis from the Bayesian perspective to undergraduate students. We will cover important concepts in Bayesian probability modeling as well as estimation using both optimization and simulation-based strategies. Key topics covered in the course include hierarchical models, mixture models, hidden Markov models and Markov Chain Monte Carlo.

Taught by: Jensen

Course usually offered in spring term

Prerequisites: A course in probability (STAT 430 or equivalent); a course in statistical inference (STAT 102, STAT 112, STAT 431 or equivalent); and experience with the statistical software R (at the level of STAT 405 or STAT 470)

Activity: Lecture

1 Course Unit

STAT 451 Fundamentals of Actuarial Science I

This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums.

Taught by: Lemaire

Course usually offered in fall term

Prerequisites: MATH 104, STAT 430. STAT 430 can be taken concurrently with BEPP 451 or STAT 451

Activity: Lecture

1 Course Unit

STAT 452 Fundamentals of Actuarial Science II

This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans.

Taught by: Lemaire

Course usually offered in spring term

Prerequisite: BEPP 451 or STAT 451

Activity: Lecture

1 Course Unit

STAT 453 Actuarial Statistics

This course covers models for insurer's losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains.

Taught by: Lemaire

Course usually offered in fall term

Prerequisite: STAT 430

Activity: Lecture

1 Course Unit

STAT 454 Applied Statistical Methods for Actuaries

One half of the course is devoted to the study of time series, including ARIMA modeling and forecasting. The other half studies modifications in random variables due to deductibles, co-payments, policy limits, and elements of simulation. This course is a possible entry point into the actuarial science program. The Society of Actuaries has approved STAT 854 for VEE credit on the topic of time series.

Taught by: Lemaire

Course usually offered in spring term

Prerequisites: STAT 430, STAT 431

Activity: Lecture

1 Course Unit

STAT 470 Data Analytics and Statistical Computing

This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests).

Taught by: Buja

One-term course offered either term

Prerequisites: STAT 101 and 102 or STAT 111 and 112 or STAT 431 or ECON 103 and ECON 104

Activity: Lecture

1 Course Unit

STAT 471 Modern Data Mining

Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class.

Taught by: Zhao

One-term course offered either term

Prerequisite: STAT 102 or 112 or 431

Activity: Lecture

1 Course Unit

STAT 474 Modern Regression for the Social, Behavioral and Biological Sciences

Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis.

Taught by: Berk

Course usually offered in spring term

Prerequisite: STAT 102 or 112 or equivalent

Activity: Lecture

1 Course Unit

STAT 475 Sample Survey Design

This course will cover the design and analysis of sample surveys. Topics include simple sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias.

Course not offered every year

Prerequisite: STAT 102 or 112 or 431

Activity: Lecture

1 Course Unit

STAT 476 Applied Probability Models in Marketing

This course will expose students to the theoretical and empirical "building blocks" that will allow them to construct, estimate, and interpret powerful models of customer behavior. Over the years, researchers and practitioners have used these models for a wide variety of applications, such as new product sales, forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.

Taught by: Fader

Course usually offered in spring term

Prerequisites: A high comfort level with basic integral calculus and recent exposure to a formal course in probability and statistics such as STAT 430 is strongly recommended.

Activity: Lecture

1 Course Unit

STAT 480 Advanced Statistical Computing

This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation.

Taught by: Buja

Course usually offered in spring term

Prerequisites: STAT 470 or STAT 405 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.

Activity: Lecture

1 Course Unit

STAT 490 Causal Inference

Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously. The course is intended for statistics concentrators and minors.

Taught by: Small

Course usually offered in spring term

Prerequisites: STAT 430 is a required course for this class. One of STAT 102, STAT 112 or STAT 431 is also required for this class.

Knowledge of R such as that covered in STAT 405 or STAT 470.

Activity: Lecture

1 Course Unit

STAT 500 Applied Regression and Analysis of Variance

An applied graduate level course in multiple regression and analysis of variance for students who have completed an undergraduate course in basic statistical methods. Emphasis is on practical methods of data analysis and their interpretation. Covers model building, general linear hypothesis, residual analysis, leverage and influence, one-way anova, two-way anova, factorial anova. Primarily for doctoral students in the managerial, behavioral, social and health sciences.

Taught by: Rosenbaum

Course usually offered in fall term

Prerequisite: STAT 102 or 112 or equivalent

Activity: Lecture

1 Course Unit

STAT 501 Introduction to Nonparametric Methods and Log-linear Models

An applied graduate level course for students who have completed an undergraduate course in basic statistical methods. Covers two unrelated topics: loglinear and logit models for discrete data and nonparametric methods for nonnormal data. Emphasis is on practical methods of data analysis and their interpretation. Primarily for doctoral students in the managerial, behavioral, social and health sciences. May be taken before STAT 500 with permission of instructor.

Taught by: Rosenbaum

Course usually offered in spring term

Prerequisite: STAT 102 or 112 or equivalent

Activity: Lecture

1 Course Unit

STAT 503 Data Analytics and Statistical Computing

This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests).

Taught by: Buja

One-term course offered either term

Prerequisites: Two courses at the statistics 400 or 500 level.

Activity: Lecture

1 Course Unit

STAT 510 Probability

Elements of matrix algebra. Discrete and continuous random variables and their distributions. Moments and moment generating functions. Joint distributions. Functions and transformations of random variables. Law of large numbers and the central limit theorem. Point estimation: sufficiency, maximum likelihood, minimum variance. Confidence intervals. One-term course offered either term

Prerequisite: A one year course in calculus

Activity: Lecture

1 Course Unit

STAT 511 Statistical Inference

Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course.

One-term course offered either term

Prerequisite: STAT 510 or equivalent

Activity: Lecture

1 Course Unit

STAT 512 Mathematical Statistics

An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures.

Course usually offered in spring term

Prerequisite: STAT 430 or 510 or equivalent

Activity: Lecture

1 Course Unit

STAT 515 Advanced Statistical Inference I

STAT 515 is aimed at first-year Ph.D. students and builds a good foundation in statistical inference from the first principles of probability.

Taught by: Low

Course usually offered in fall term

Prerequisites: STAT 430 and STAT 431 and MATH 114 and MATH 240 or equivalent

Activity: Lecture

1 Course Unit

STAT 516 Advanced Statistical Inference II

STAT 516 is a natural continuation of STAT 515, and the main focus is on asymptotic evaluations and regression models. Time permitting, it also discusses some basic nonparametric statistical methods.

Taught by: Ma

Course usually offered in spring term

Prerequisite: STAT 515

Activity: Lecture

1 Course Unit

STAT 520 Applied Econometrics I

This is a course in econometrics for graduate students. The goal is to prepare students for empirical research by studying econometric methodology and its theoretical foundations. Students taking the course should be familiar with elementary statistical methodology and basic linear algebra, and should have some programming experience. Topics include conditional expectation and linear projection, asymptotic statistical theory, ordinary least squares estimation, the bootstrap and jackknife, instrumental variables and two-stage least squares, specification tests, systems of equations, generalized least squares, and introduction to use of linear panel data models.

Taught by: Shaman

Course usually offered in fall term

Prerequisites: MATH 114 and MATH 312 or equivalents, and an undergraduate introduction to probability and statistics

Activity: Lecture

1 Course Unit

STAT 521 Applied Econometrics II

Topics include system estimation with instrumental variables, fixed effects and random effects estimation, M-estimation, nonlinear regression, quantile regression, maximum likelihood estimation, generalized method of moments estimation, minimum distance estimation, and binary and multinomial response models. Both theory and applications will be stressed.

Taught by: Shaman

Course usually offered in spring term

Prerequisites: STAT 520. This is a continuation of STAT 520

Activity: Lecture

1 Course Unit

STAT 533 Stochastic Processes

An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes.

One-term course offered either term

Prerequisite: STAT 510 or equivalent

Activity: Lecture

1 Course Unit

STAT 542 Bayesian Methods and Computation

Sophisticated tools for probability modeling and data analysis from the Bayesian perspective. Hierarchical models, mixture models and Monte Carlo simulation techniques.

Taught by: Jensen

Course usually offered in spring term

Prerequisite: STAT 430 or 510 or equivalent or permission of instructor

Activity: Lecture

1 Course Unit

STAT 571 Modern Data Mining

Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class.

Taught by: Zhao

One-term course offered either term

Prerequisite: Two courses at the statistics 400 or 500 level or permission from instructor

Activity: Lecture

1 Course Unit

STAT 580 Advanced Statistical Computing

This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation.

Taught by: Buja

Course usually offered in spring term

Prerequisites: STAT 503 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.

Activity: Lecture

1 Course Unit

STAT 590 Causal Inference

Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously.

Taught by: Small

Course usually offered in spring term

Prerequisites: Background in probability and statistics; some knowledge of R.

Activity: Lecture

1 Course Unit

STAT 613 Regression Analysis for Business

This course provides the fundamental methods of statistical analysis, the art and science of extracting information from data. The course will begin with a focus on the basic elements of exploratory data analysis, probability theory and statistical inference. With this as a foundation, it will proceed to explore the use of the key statistical methodology known as regression analysis for solving business problems, such as the prediction of future sales and the response of the market to price changes. The use of regression diagnostics and various graphical displays supplement the basic numerical summaries and provides insight into the validity of the models. Specific important topics covered include least squares estimation, residuals and outliers, tests and confidence intervals, correlation and autocorrelation, collinearity, and randomization. The presentation relies upon computer software for most of the needed calculations, and the resulting style focuses on construction of models, interpretation of results, and critical evaluation of assumptions.

Course usually offered in fall term

Prerequisites: The basic mathematical skills covered in STAT 611, Mathematics for Business Analysis

Activity: Lecture

1 Course Unit

Notes: Lecture and discussion, assigned exercises, data analysis project, quizzes and a final exam.

STAT 621 Accelerated Regression Analysis for Business

STAT 621 is intended for students with recent, practical knowledge of the use of regression analysis in the context of business applications. This course covers the material of STAT 613, but omits the foundations to focus on regression modeling. The course reviews statistical hypothesis testing and confidence intervals for the sake of standardizing terminology and introducing software, and then moves into regression modeling. The pace presumes recent exposure to both the theory and practice of regression and will not be accommodating to students who have not seen or used these methods previously. The interpretation of regression models within the context of applications will be stressed, presuming knowledge of the underlying assumptions and derivations. The scope of regression modeling that is covered includes multiple regression analysis with categorical effects, regression diagnostic procedures, interactions, and time series structure. The presentation of the course relies on computer software that will be introduced in the initial lectures.

Taught by: George

Course usually offered in fall term

Prerequisites: Recent exposure to the theory and practice of regression modeling.

Activity: Lecture

0.5 Course Units

Notes: Lecture and discussion, assigned exercises, data analysis, quizzes, and a final exam.

STAT 701 Modern Data Mining

Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class.

Taught by: Zhao

One-term course offered either term

Prerequisite: STAT 613 or equivalent

Activity: Lecture

1 Course Unit

STAT 705 Statistical Computing with R

The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives.

Taught by: Stine, Waterman, Zhang

One-term course offered either term

Prerequisites: STAT 613 or STAT 621 or waiving the Statistics Core completely.

Activity: Lecture

0.5 Course Units

STAT 711 Forecasting Methods for Management

This course provides an introduction to the wide range of techniques available for statistical forecasting. Qualitative techniques, smoothing and decomposition of time series, regression, adaptive methods, autoregressive-moving average modeling, and ARCH and GARCH formulations will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations.

Taught by: Shaman

One-term course offered either term

Prerequisite: STAT 613 or equivalent

Activity: Lecture

1 Course Unit

STAT 722 Predictive Analytics for Business (formerly STAT 622)

This course follows from the introductory regression classes, STAT 102, STAT 112, and STAT 431 for undergraduates and STAT 613 for MBAs.

It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest.

One-term course offered either term

Prerequisite: STAT 613 or STAT 621 or having waived the statistics core completely

Activity: Lecture

0.5 Course Units

STAT 724 Text Analytics

This course introduces methods for the analysis of unstructured data, focusing on statistical models for text. Techniques include those for sentiment analysis, topic models, and predictive analytics. Course includes topics from natural language processing (NLP), such as identifying parts of speech, parsing sentences (e.g., subject and predicate), and named entity recognition (people and places). Unsupervised techniques suited to feature creation provide variables suited to traditional statistical models (regression) and more recent approaches (regression trees). Examples that span the course illustrate the success of text analytics. Hierarchical generating models often associated with nonparametric Bayesian analysis supply theoretical foundations.

Taught by: Stine

One-term course offered either term

Prerequisites: Students should be familiar with regression models at the level of STAT 613 and the R statistics language at the level of STAT 705. Familiarity with the R-Studio development environment is presumed, as well as common R packages such as stringr, dplyr and ggplot. Those with more knowledge of Statistics, such as from STAT 722, or computing skills will benefit. The predominant software used in the course is R, with bits of JMP when helpful for interactive illustration. Familiarity with basic probability models is helpful but not presumed.

Activity: Lecture

0.5 Course Units

STAT 770 Data Analytics and Statistical Computing

This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests).

Taught by: Buja

One-term course offered either term

Prerequisites: STAT 613 or STAT 621 or waiving the Statistics Core completely.

Activity: Lecture

1 Course Unit

STAT 776 Applied Probability Models in Marketing

This course will expose students to the theoretical and empirical "building blocks" that will allow them to develop and implement powerful models of customer behavior. Over the years, researchers and practitioners have used these methods for a wide variety of applications, such as new product sales forecasting, analyses of media usage, customer valuation, and targeted marketing programs. These same techniques are also very useful for other types of business (and non-business) problems. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.

Taught by: Fader

Course usually offered in spring term

Prerequisites: Students must have a high comfort level with basic integral calculus, and recent exposure to a formal course in probability and statistics is strongly recommended.

Activity: Lecture

1 Course Unit

Notes: Format: Lecture, real-time problem solving

STAT 780 Advanced Statistical Computing

This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation.

Taught by: Buja

Course usually offered in spring term

Prerequisites: STAT 770 or STAT 705 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.

Activity: Lecture

1 Course Unit

STAT 851 Fundamentals of Actuarial Science I

This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums.

Taught by: Lemaire

Course usually offered in fall term

Prerequisite: One semester of calculus

Activity: Lecture

1 Course Unit

STAT 852 Fundamentals of Actuarial Science II

This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans.

Taught by: Lemaire

Course usually offered in spring term

Prerequisite: STAT 851 or BEPP 851

Activity: Lecture

1 Course Unit

STAT 853 Actuarial Statistics

This course covers models for insurer's losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains.

Taught by: Lemaire

Course usually offered in fall term

Prerequisite: Two semesters of Statistics

Activity: Lecture

1 Course Unit

STAT 854 Applied Statistical Methods for Actuaries

One half of the course is devoted to the study of time series, including ARIMA modeling and forecasting. The other half studies modifications in random variables due to deductibles, co-payments, policy limits, and elements of simulation. This course is a possible entry point into the actuarial science program. The Society of Actuaries has approved STAT 854 for VEE credit on the topic of time series.

Taught by: Lemaire

Course usually offered in spring term

Prerequisite: One semester of probability

Activity: Lecture

1 Course Unit

STAT 899 Independent Study

One-term course offered either term

Prerequisites: Written permission of instructor, the department MBA advisor and course coordinator.

Activity: Independent Study

1 Course Unit

STAT 910 Forecasting and Time Series Analysis

Fourier analysis of data, stationary time series, properties of autoregressive moving average models and estimation of their parameters, spectral analysis, forecasting. Discussion of applications to problems in economics, engineering, physical science, and life science.

Taught by: Stine

Course offered spring; odd-numbered years

Prerequisite: STAT 520 or 961 or equivalent

Activity: Lecture

1 Course Unit

STAT 915 Nonparametric Inference

Statistical inference when the functional form of the distribution is not specified. Nonparametric function estimation, density estimation, survival analysis, contingency tables, association, and efficiency.

Course not offered every year

Prerequisite: STAT 520 or equivalent

Activity: Lecture

1 Course Unit

STAT 920 Sample Survey Methods

This course will cover the design and analysis of sample surveys. Topics include simple random sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias.

Taught by: Small

Course not offered every year

Prerequisites: STAT 520, 961 or 970 or permission of instructor

Activity: Lecture

1 Course Unit

STAT 921 Observational Studies

This course will cover statistical methods for the design and analysis of observational studies. Topics will include the potential outcomes framework for causal inference; randomized experiments; matching and propensity score methods for controlling confounding in observational studies; tests of hidden bias; sensitivity analysis; and instrumental variables.

Taught by: Small

Course usually offered in fall term

Prerequisites: STAT 520, 961 or 970 or permission of instructor

Activity: Lecture

1 Course Unit

STAT 925 Multivariate Analysis: Theory

This is a course that prepares PhD students in statistics for research in multivariate statistics and high dimensional statistical inference. Topics from classical multivariate statistics include the multivariate normal distribution and the Wishart distribution; estimation and hypothesis testing of mean vectors and covariance matrices; principal component analysis, canonical correlation analysis and discriminant analysis; etc. Topics from modern multivariate statistics include the Marcenko-Pastur law, the Tracy-Widom law, nonparametric estimation and hypothesis testing of high-dimensional covariance matrices, high-dimensional principal component analysis, etc.

Taught by: Ma

Course not offered every year

Prerequisites: STAT 930, 970 and 972 or permission of instructor

Activity: Lecture

1 Course Unit

STAT 926 Multivariate Analysis: Methodology

This is a course that prepares PhD students in statistics for research in multivariate statistics and data visualization. The emphasis will be on a deep conceptual understanding of multivariate methods to the point where students will propose variations and extensions to existing methods or whole new approaches to problems previously solved by classical methods. Topics include: principal component analysis, canonical correlation analysis, generalized canonical analysis; nonlinear extensions of multivariate methods based on optimal transformations of quantitative variables and optimal scaling of categorical variables; shrinkage- and sparsity-based extensions to classical methods; clustering methods of the k-means and hierarchical varieties; multidimensional scaling, graph drawing, and manifold estimation.

Taught by: Buja

Course not offered every year

Prerequisite: STAT 961 or permission of instructor

Activity: Lecture

1 Course Unit

STAT 927 Bayesian Statistical Theory and Methods

This graduate course will cover the modeling and computation required to perform advanced data analysis from the Bayesian perspective. We will cover fundamental topics in Bayesian probability modeling and implementation, including recent advances in both optimization and simulation-based estimation strategies. Key topics covered in the course include hierarchical and mixture models, Markov Chain Monte Carlo, hidden Markov and dynamic linear models, tree models, Gaussian processes and nonparametric Bayesian strategies.

Taught by: Jensen

Course not offered every year

Prerequisite: STAT 430 or STAT 510

Activity: Lecture

1 Course Unit

STAT 928 Statistical Learning Theory

Statistical learning theory studies the statistical aspects of machine learning and automated reasoning, through the use of (sampled) data. In particular, the focus is on characterizing the generalization ability of learning algorithms in terms of how well they perform on "new" data when trained on some given data set. The focus of the course is on: providing the fundamental tools used in this analysis; understanding the performance of widely used learning algorithms; understanding the "art" of designing good algorithms, both in terms of statistical and computational properties. Potential topics include: empirical process theory; online learning; stochastic optimization; margin based algorithms; feature selection; concentration of measure.

Course usually offered in spring term

Prerequisites: Probability and linear algebra.

Activity: Lecture

1 Course Unit

STAT 930 Probability

Measure theory and foundations of Probability theory. Zero-one Laws. Probability inequalities. Weak and strong laws of large numbers. Central limit theorems and the use of characteristic functions. Rates of convergence. Introduction to Martingales and random walk.

Taught by: Pemantle

Course usually offered in fall term

Prerequisite: STAT 430 or 510 or equivalent

Activity: Lecture

1 Course Unit

STAT 931 Stochastic Processes

Markov chains, Markov processes, and their limit theory. Renewal theory. Martingales and optimal stopping. Stable laws and processes with independent increments. Brownian motion and the theory of weak convergence. Point processes.

Taught by: Pemantle

Course usually offered in spring term

Prerequisite: STAT 930

Activity: Lecture

1 Course Unit

STAT 955 Stochastic Calculus and Financial Applications

Selected topics in the theory of probability and stochastic processes. Course usually offered in fall term

Prerequisite: STAT 930 or equivalent

Activity: Lecture

1 Course Unit

STAT 957 Seminar in Data Analysis

Survey of methods for the analysis of large unstructured data sets: detection of outliers, Winsorizing, graphical techniques, robust estimators, multivariate problems.

Course not offered every year

Prerequisites: STAT 961, 971, 972, 925, or equivalents; permission of instructor

Activity: Seminar

1 Course Unit

STAT 961 Statistical Methodology

This is a course that prepares 1st year PhD students in statistics for a research career. This is not an applied statistics course. Topics covered include: linear models and their high-dimensional geometry, statistical inference illustrated with linear models, diagnostics for linear models, bootstrap and permutation inference, principal component analysis, smoothing and cross-validation.

Taught by: Buja

Course usually offered in fall term

Prerequisites: STAT 431 or 520 or equivalent; a solid course in linear algebra and a programming language

Activity: Lecture

1 Course Unit

STAT 962 Advanced Methods for Applied Statistics

This course is designed for Ph.D. students in statistics and will cover various advanced methods and models that are useful in applied statistics. Topics for the course will include missing data, measurement error, nonlinear and generalized linear regression models, survival analysis, experimental design, longitudinal studies, building R packages and reproducible research.

Taught by: Small

Course usually offered in spring term

Prerequisite: STAT 961

Activity: Lecture

1 Course Unit

STAT 970 Mathematical Statistics

Decision theory and statistical optimality criteria, sufficiency, point estimation and hypothesis testing methods and theory.

Taught by: Small

Course usually offered in fall term

Prerequisites: STAT 431 or 520 or equivalent; comfort with mathematical proofs (e.g., MATH 360)

Activity: Lecture

1 Course Unit

STAT 971 Introduction to Linear Statistical Models

Theory of the Gaussian Linear Model, with applications to illustrate and complement the theory. Distribution theory of standard tests and estimates in multiple regression and ANOVA models. Model selection and its consequences. Random effects, Bayes, empirical Bayes and minimax estimation for such models. Generalized (Log-linear) models for specific non-Gaussian settings.

Taught by: Ma

Course usually offered in spring term

Prerequisite: STAT 970

Activity: Lecture

1 Course Unit

STAT 972 Advanced Topics in Mathematical Statistics

A continuation of STAT 970.

Taught by: Cai

One-term course offered either term

Prerequisites: STAT 970 and 971

Activity: Lecture

1 Course Unit

STAT 974 Modern Regression for the Social, Behavioral and Biological Sciences

Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis.

Taught by: Berk

Course usually offered in spring term

Prerequisites: Two statistics courses at the graduate school level including a solid foundation in the generalized linear model.

Activity: Lecture

1 Course Unit

STAT 991 Seminar in Advanced Application of Statistics

This seminar will be taken by doctoral candidates after the completion of most of their coursework. Topics vary from year to year and are chosen from advance probability, statistical inference, robust methods, and decision theory with principal emphasis on applications.

One-term course offered either term

Activity: Seminar

1 Course Unit

STAT 995 Dissertation

One-term course offered either term

Activity: Dissertation

1 Course Unit

STAT 999 Independent Study

One-term course offered either term

Prerequisites: Written permission of instructor and the department course coordinator.

Activity: Independent Study

1 Course Unit